

PCT/KR 20/539762
03/02569
PCT/KR 26.11.2003



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

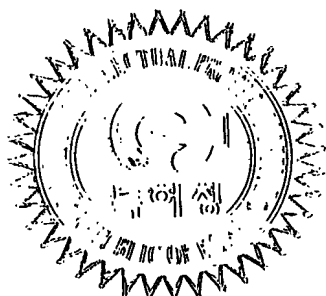
This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원 번호 : 10-2002-0074009
Application Number

출원 년 월 일 : 2002년 11월 26일
Date of Application NOV 26, 2002

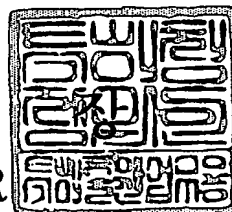
PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

출원인 : 한국전자통신연구원
Applicant(s) Electronics and Telecommunications Research Ins



2003 년 11 월 26 일

특 허 청
COMMISSIONER



【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【제출일자】	2002.11.26
【발명의 명칭】	엘리먼트 기반 웹 문서의 파싱시스템 및 그 방법
【발명의 영문명칭】	Parsing system and method of Multi-document based on elements
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	권태복
【대리인코드】	9-2001-000347-1
【포괄위임등록번호】	2001-057650-1
【대리인】	
【성명】	이화익
【대리인코드】	9-1998-000417-9
【포괄위임등록번호】	1999-021997-1
【발명자】	
【성명의 국문표기】	최은정
【성명의 영문표기】	CHOI, Eun Jeong
【주민등록번호】	760903-2711011
【우편번호】	780-945
【주소】	경상북도 경주시 성건동 719-8번지
【국적】	KR
【발명자】	
【성명의 국문표기】	손지연
【성명의 영문표기】	SON, Ji Yeon
【주민등록번호】	681205-2177815
【우편번호】	305-707
【주소】	대전광역시 유성구 신성동 삼성한울아파트 108동 405호
【국적】	KR

【발명자】

【성명의 국문표기】 한동원
【성명의 영문표기】 HAN, Dong Won
【주민등록번호】 580916-1074415
【우편번호】 302-280
【주소】 대전광역시 서구 월평동 누리아파트 107동1008호
【국적】 KR

【공지예외적용대상증명서류의 내용】

【공개형태】 학술단체 서면발표
【공개일자】 2002.10.25
【심사청구】 청구
【취지】 특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인
 권태복 (인) 대리인
 이화익 (인)

【수수료】

【기본출원료】	20 면	29,000 원
【가산출원료】	0 면	0 원
【우선권주장료】	0 건	0 원
【심사청구료】	8 항	365,000 원
【합계】		394,000 원
【감면사유】	정부출연연구기관	
【감면후 수수료】		197,000 원

【기술이전】

【기술양도】 희망
【실시권 허여】 희망
【기술지도】 희망

【첨부서류】

1. 공지예외적용대상(신규성상실의예외, 출원시의특례)규정을 적용받기 위한 증명서류_1통 2. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 기존의 유무선 웹사이트들을 그대로 활용하여 이를 휴대 단말에서 브라우징하기 위해, 단일 혹은 다중 마크업으로 구성된 웹 문서를 종류에 상관없이 파싱하는 웹 문서 파싱시스템 및 방법에 관한 것이다.

본 발명의 웹 문서 파싱시스템은 지원하고자 하는 문서종류에서 필요로 하는 모든 마크업 데이터를 위한 토큰 테이블을 참조하여 마크업과, non-마크업을 기반으로 토큰을 분리 생성하는 어휘 분석기와, 각 문서의 DTD를 기반으로 콘텐츠 모델을 분석하여 이를 기반으로 각 구문을 분석한 후, 상기 단말의 그래픽 사용자 인터페이스(GUI)를 기반으로 트리 기반의 객체를 형성하는 구문 분석기를 포함하여 웹 서버로부터 다양한 웹 문서를 호출하여 휴대 단말의 응용에 제공한다.

본 발명에 의하면, 기존에 구축된 웹 사이트를 그대로 이용할 수 있을 뿐만 아니라, 단말의 응용 프로그램이 필요로 하는 정보만을 추출하는 것이 가능하므로, 인터넷 서비스 제공자는 각 단말에 특화된 웹 사이트를 구축하지 않아도 되어, 웹 사이트 구축에 따른 시간적, 금전적 비용을 감소시킬 수 있다.

【대표도】

도 3

【색인어】

파서, 어휘분석기, 구문분석기, 웹 서버, 마크업언어, 엘리먼트

0074009

출력 일자: 2003/12/3

【명세서】

【발명의 명칭】

엘리먼트 기반 웹 문서의 파싱시스템 및 그 방법{ Parsing system and method of
Multi-document based on elements }

【도면의 간단한 설명】

도 1은 종래의 방식에 따라 휴대 단말기에 웹 문서를 표시하는 구성을 도시한 개략도,
도 2는 본 발명에 따른 웹 문서 파싱 시스템을 이용하여 휴대 단말기에 웹 문서를 표시
하는 구성을 도시한 블록도,

도 3은 본 발명에 따른 웹 문서 파싱 시스템의 구성을 도시한 개략도,
도 4는 도 3에 도시된 어휘분석기의 동작을 설명하기 위해 도시한 개념도,
도 5는 본 발명에 따른 문법구조의 예를 도시한 개념도,
도 6은 본 발명에 따른 통합 파서의 파싱 절차를 도시한 순서도.

*도면의 주요부분에 대한 부호의 설명

210;휴대 단말기 212;응용

214;통합 파서 230;웹 서버

310;어휘분석기 311;토큰테이블

312;토큰발생부 313;XML Well-formedness 검사부

320;구문분석기 321;문서 컨텐츠 모델

322;XML검증부 323;GUI기반 객체 생성부

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

- <14> 본 발명은 웹 문서를 휴대용 단말기에 브라우징하기 위한 파서(parser)에 관한 것으로, 상세하게는 다양한 종류의 마크업 언어로 작성된 웹 문서들을 통합하여 지원할 수 있는 웹 문서 통합 파싱 시스템 및 그 방법에 관한 것이다.
- <15> 도 1은 종래의 방식에 따라 휴대 단말기에 웹 문서를 브라우징하는 구성을 도시한 개략도이다.
- <16> 도면을 참조하면 웹 서버(130)는 다양한 마크업 언어로 된 웹 문서들을 각각 구비하고 있고, 휴대용 단말기(110)는 각 마크업 언어를 지원하는 브라우저들 예를 들면, HDML 브라우저(111), WML 브라우저(112) 및 mHTML 브라우저(113) 등을 각각 탑재한 후 WAP 게이트웨이(120)를 통하거나 직접 웹 서버(130)에 접속하여 해당 웹 문서를 브라우징하도록 되어 있다.
- <17> 이러한 구성에 의하면, 하나의 단말로 다양한 문서를 보려면 지원하는 마크업 언어의 수에 해당하는 다수의 브라우저를 각각 탑재하여 해당 언어의 문서를 브라우징해야 하므로 휴대 단말의 구성이 복잡해지는 문제점이 있다.
- <18> 따라서 근래에는 휴대폰의 사용이 일반화되면서 무선 인터넷 서비스를 지원하기 위해 기존의 HTML(Hyper Text Markup Language)로부터 파생된 마크업 언어들이 등장하게 되었다.
- <19> 기존의 HTML을 사용해 무선 인터넷을 서비스하지 않고, 다른 마크업 언어를 개발하게 된 이유로는 크게 무선 선로에 따른 제약과, 휴대 단말에 따른 제약을 들 수 있다. 현재의 휴대폰과 같은 이동 단말기에서 무선선로는 유선에 비해 상대적으로 대역폭이 좁고 에러 발생률이

높은 제약이 있고, 휴대 단말기 자체는 유선 인터넷에서 사용되는 데스크 탑 컴퓨터에 비해 화면 창의 크기(Window size)가 작고, CPU(Central Processing Unit)와 메모리 등의 컴퓨팅 능력이 데스크 탑 PC (Personal Computer)에 비해 상대적으로 열등하다. 반면, 기존의 유선 인터넷에서 제공되는 HTML은 기능이 많고 처리과정이 복잡하여 휴대 단말에서 지원하기 어려웠던 것이다.

20> 이러한 점을 감안하여, HTML의 기능 중 일부를 그대로 상속하는 한편 각 단말에 특화된 마크업 언어를 개발하게 되었고, 대표적인 예로써, HDML(Handheld Device Markup Language), WML(Wireless Markup Language), mHTML(Mobile HTML), cHTML(Compact HTML) 등이 등장하여 서비스되고 있다.

21> 그러나 상기한 각 마크업 언어들은 서비스 제공자 및 단말의 특성을 고려하여 개별적으로 개발된 언어들로서, 서로 호환성이 되지 않는 문제가 있다. 즉, 인터넷 서비스 제공자가 두 가지 종류의 단말에 같은 내용의 콘텐츠를 제공하고자 한다면, 각 단말에서 처리 가능한 마크업 규칙을 따르도록 두개의 콘텐츠를 개발해야 하는 문제점이 있고, 단말 사용자 입장에서는 다른 인터넷 서비스 제공자가 제공하는 콘텐츠를 볼 수 없게 되는 문제점이 있다.

【발명이 이루고자 하는 기술적 과제】

22> 본 발명은 이러한 문제점을 해결하기 위한 것으로, 그 목적은 기존에 구축된 유,무선 웹사이트에서 제공하는 다양한 마크업 언어로 작성된 콘텐츠들을 휴대용 단말의 사양에 제한없이 통합하여 브라우징할 수 있는 엘리먼트 기반 웹 문서의 파싱 시스템 및 방법을 제공하는 것이다.

- <23> 본 발명의 다른 목적은 종류가 다른 마크업 언어들의 특성을 분석하여 XML에서 정의하는 엘리먼트(element)를 기반으로 문서를 파싱하는 한편, 단말에서 처리 가능한 엘리먼트만을 선별하여 데이터로 저장할 수 있도록 함으로써, 인터넷 서비스 대역을 확장할 수 있는 엘리먼트 기반 웹 문서의 파싱 시스템 및 방법을 제공하는 것이다.
- <24> 상기 목적을 달성하기 위한 본 발명의 엘리먼트 기반 웹 문서의 파싱 시스템은 웹 서버로부터 다양한 웹 문서를 호출하여 휴대 단말의 응용에 제공하기 위한 휴대 단말의 웹 문서 파싱 시스템에 있어서, 지원하고자 하는 문서종류에서 필요로 하는 모든 마크업 데이터를 위한 토큰 테이블을 참조하여 마크업; non-마크업을 기반으로 토큰을 분리생성하는 어휘 분석기; 및 각 문서의 DTD를 기반으로 콘텐츠 모델을 분석하여 이를 기반으로 각 구문을 분석한 후, 상기 단말의 그래픽 사용자 인터페이스(GUI)를 기반으로 트리 기반의 객체를 형성하는 구문 분석기;를 포함한다.
- <25> 상기 어휘 분석기는 주석이나 공백을 처리하는 주석문 분석기, 마크업 시작태그를 인식하여 토큰을 생성하는 마크업 시작 분석기, 속성을 분석하여 토큰을 생성하는 속성 분석기 및 일반 사용자 데이터를 분석하여 토큰을 생성하는 일반 사용자 데이터 분석기를 포함하고, 상기 구문 분석기는 상기 어휘 분석기에서 생성된 토큰을 바탕으로 해당 문서가 각각의 DTD에 적합하게 작성되었는지 여부를 검사하는 XML 검증부 및 상기 분석된 마크업을 단말의 그래픽 사용자 인터페이스와 일치시키는 단말 GUI기반 객체 생성부로 이루어진다.
- <26> 또한 상기 목적을 달성하기 위한 본 발명의 웹 서버로부터 다양한 웹 문서를 호출하여 휴대 단말의 응용에 제공하기 위한 엘리먼트 기반 웹 문서의 파싱방법은 웹 서버로부터 호출한 웹 문서를 파싱하기 위한 방법에 있어서, 상기 웹 문서로부터 토큰을 읽어와 분석하는 제 1단계; 상기 토큰을 분석한 결과 정의된 시작태그가 아니거나 주석문 또는 공백일 경우 그 토큰을

무시하고, 정의된 시작태그가 읽힐 경우 상기 토큰으로부터 엘리먼트의 속성을 파싱하는 제 2 단계; 상기 토큰으로부터 엘리먼트 속성을 파싱하고 엘리먼트의 GUI관련 정보를 저장한 후 엘리먼트의 내용을 파싱하는 제 3단계; 상기 3단계에서 엘리먼트의 내용을 파싱한 결과 일반 사용자 데이터이면 내용의 GUI관련 정보를 저장하고, 일반 사용자 데이터가 아닐 경우에 마침태크가 나올 때까지 데이터를 읽는 제 4단계; 및 상기 제 4단계에서 일반 사용자 데이터가 아닌 상태에서 정의된 시작태그의 마침태크가 나타나면 종료하고, 그렇지 않을 경우 무시하고 리턴하는 제 5단계;의 루틴을 수행하는 것을 특징으로 한다.

【발명의 구성 및 작용】

<27> 이하, 본 발명의 바람직한 실시예를 첨부한 도면을 참조하여 상세하게 설명한다.

<28> 본 발명은 기존에 구축된 인터넷 상의 웹 서버에 의해 제공되는 모든 종류의 콘텐츠를 휴대 단말기의 제약 사항에 관련없이 사용자에게 서비스 가능하도록 하기 위해 웹 페이지를 호출하여 엘리먼트를 기반으로 파싱하여 추출된 정보를 응용 프로그램에 전달하도록 구성을 개시하고자 하는 것으로, 현재 서비스되고 있는 마크업 언어들을 분석해보면 다음 표 1과 같이 3가지의 종류로 분류해 볼 수 있다.

<29> 【표 1】

구분	단일 문서 구조	내장형 구조	모듈화 구조
마크업 언어	XHTML WML CHTML MHTML HTML	WML2 namespace를 이용해 다른 형식의 마크업언어를 내장하는 방법 객체태그를 이용한 객체내장 프로토콜을 이용한 객체내장	XHTML모듈화

- <30> 상기 표 1을 참조하여 분류된 마크업 언어들을 분석해보면 HTML문서를 제외한 대부분의 문서는 XML을 기반으로 개발된 것들이며, HTML 역시 XML로 전이하고 있는 추세이다. 따라서 본 발명은 XML을 기반으로 하는 마크업 언어들에 기초하여 통합 파싱시스템을 개시하고자 하는 것이다.
- <31> 도 2는 본 발명에 따른 웹 문서 파싱 시스템을 이용하여 휴대 단말기에 웹 문서를 브라우징하는 전체 구성을 도시한 블록도이다.
- <32> 도면을 참조하면, 본 발명은 인터넷을 통해 소정의 언어로 작성된 웹 문서들을 웹 서버(23)에서 제공하고, 본 발명이 적용된 휴대용 단말기(210)는 웹 서버(230)에서 제공된 문서를 파싱하는 통합파서(214)와 상기 통합파서(214)로부터 추출된 정보를 이용하는 응용 프로그램(212)을 탑재하고 있다.
- <33> 여기서, 통합 파서(214)는 웹 서버(230)로부터 가져온 문서들을 입력으로 하고, 응용 프로그램(212)에서 필요로 하는 정보를 메모리 혹은 하드 디스크에 저장한 데이터를 출력한다.
- <34> 그리고 웹 서버(230)에서 제공되는 문서는 HTML, XHTML, MHTML, CHTML, WML, HDML 등의 SGML 또는 XML을 기반으로 프리젠테이션(presentation)을 목적으로 작성된 문서를 모두 포함한다.
- <35> 도 3은 본 발명에 따른 웹 문서 파싱시스템의 구성을 도시한 개략도이고, 도 4는 도 3에 도시된 어휘 분석기의 동작을 설명하기 위해 도시한 개념도이며, 도 5는 본 발명에 따른 문법 구조의 예를 도시한 개념도이다.
- <36> 본 발명의 파싱 시스템은 도 3에 도시된 바와 같이, 크게 어휘 분석기(310)와 구문 분석기(320)로 구성된다.

- <37> 어휘 분석기(310)는 지원하고자 하는 문서 종류에서 필요로 하는 모든 마크업 데이터를 위한 토큰 테이블(311)을 참조하여 마크업과 non-마크업을 기반으로 토큰을 분리한다.
- <38> 구문 분석기(320)는 XML 검증부(322)와 GUI기반 객체 생성부(323)로 구성되어 모든 문서의 마크업 언어를 각각의 휴대 단말에 적합하게 표현할 수 있도록 도와 주는 것으로, 각 문서의 DTD를 기반으로 콘텐츠 모델(321)을 분석하여 이를 기반으로 구문을 분석한 후, 단말의 그래픽 사용자 인터페이스를 기반으로 트리 기반의 객체를 형성하여 렌더링을 위한 데이터로 제공한다.
- <39> 어휘 분석기(310)는 토큰발생부(312)와 XML Well-formedness 검사부(313)를 구비하여 XML Well-formedness 표준에 기반하여 토큰을 추출해 내며, 이때 지원하고자 하는 문서들의 모든 토큰을 테이블(311)로 구성한다. 그리고 도 4에 도시된 바와 같이 XML 구조에 따라 상태변이를 하면서 토큰을 분리하게 된다.
- <40> 여기서, 토큰의 상태란 같은 어휘라 할지라도 어휘 분석기 상태에 따라 다른 토큰으로 분리해 내도록 하는 것을 말한다. 어휘 분석기(310)의 상태는 주석문(comment), 마크업 시작(start tag), 속성(attrStart, attValue) 및 일반 사용자 데이터(pc data) 등으로 나뉜다.
- <41> 도 4를 참조하면, 일반적으로 웹 문서는 초기상태일 때 공백이나 시작(Start)태그, 종료(End)태그 등을 포함하게 되는데, 본 발명의 어휘 분석기(310)는 주석분석기(410), 마크업 시작분석기(420), 속성분석기(430,440) 및 일반 사용자 데이터 분석기(450)를 통해 웹 문서를 각각 분석함으로써 토큰(Token)을 생성하게 된다.
- <42> 상기 어휘 분석기(310)에서는 XML Well-formedness 표준을 기반으로 문서를 분석하여 토큰을 추출하고, 구문 분석기(320)에서는 어휘 분석기(310)에서 추출된 토큰을 이용하여 문서가

각각의 DTD(Document Type Definition)에 적합하게 작성되었는지 여부를 검사하게 되며, 이러한 과정에서 분석된 마크업은 단말의 그래픽 사용자 인터페이스(GUI)와 일치시킨다.

- <43> 즉, 어휘 분석기(310)는 마크업 언어의 그래픽 사용자 인터페이스 모델(321)을 휴대 단말에서 지원 가능한 그래픽 사용자 인터페이스(GUI)로 표현할 수 있도록 하기 위해 매핑(mapping)을 하는 것이다.
- <44> 이와 같이 매핑하는 이유는 휴대 단말들은 각각의 특성에 맞는 그래픽 사용자 인터페이스를 가지므로, 모든 마크업 언어 표준들을 데스크 탑에서와 같이 지원할 수는 없고, 따라서 마크업 언어의 그래픽 사용자 인터페이스 특성들을 해당 휴대 단말의 그래픽 사용자 인터페이스에 맞게 고쳐야 하기 때문이다.
- <45> 본 발명의 구문 분석기(320)는 여러 종류의 문서 혹은 다중문서를 파싱하기 위해 도 5에 도시된 바와 같은 문법구조를 정의한다.
- <46> 즉, 도 5의 문법구조에서 파서는 여러 종류의 표준을 지원하는 마크업 언어를 파싱할 수 있는 것으로, 지원하고자 하는 모든 DTD를 분석하여 각 요소별로 문법을 정의한다.
- <47> 이때, 각 요소나 속성들을 살펴보면, 대부분의 요소나 속성들은 여러 언어에서 공통으로 사용될 수 있으나 어떤 요소나 속성들은 특정 언어에 국한되므로, 본 발명에서는 모든 프리젠테이션을 위한 마크업들의 공통 분모를 파싱할 수 있도록 시스템을 설계하고 있다.
- <48> 도 5와 같은 문법구조를 BNF 형식으로 표현하면 다음 표 2와 같다.

<49> 【표 2】

[1] document: Language A Language B Language C
[2] Language A: [Element A' Element B']* Language B Language C....
[3] Element A': attributes contents

- 50> 상기 표 2의 문법을 설명하면, 라인[1]은 파싱하고자 하는 하나의 문서에 대해 여러 표준을 지원하는 언어들 중의 하나로 이루어진다. 라인[2]에서는 각각의 언어에 대해 자신의 DTD를 기반으로 구성된 콘텐츠 모델을 가짐과 동시에 다른 언어를 내장할 수도 있다. 라인[3]~[5]의 각 요소는 속성들과 자신의 콘텐츠들을 가질 수 있다. 라인 [6]은 라인[2]와 마찬가지로 다른 표준을 지원하는 콘텐츠 역시 자신의 DTD를 기반으로 한 콘텐츠 모델과 다른 언어를 내장할 수 있음을 나타낸다.
- 51> 이하, 도 6을 참조하여 상기와 같이 구성된 본 발명에 따른 웹 문서 파싱 시스템이 다양한 웹 문서를 엘리먼트를 기반으로 파싱하는 파싱과정을 설명한다.
- 52> 도면을 참조하면, 본 발명의 통합 파서기(214)는 파싱의 시작과 끝을 최상위 엘리먼트를 인식하여 상기 엘리먼트의 시작태그를 인식하면서 시작되어 상기 엘리먼트의 마침태그를 인식하여 종료한다.
- 53> 본 발명은 어휘 분석기(310)에서 요청에 응답한 웹 문서를 분석하여 생성된 토큰을 읽어와 주석문이나 공백인지를 판단하고, 만약 읽어온 토큰이 주석문이나 공백일 경우 모두 읽어들이거나 이들 토큰들에 대한 처리를 하지 않고, 다시 엘리먼트를 인식하기 위해 토큰을 읽어들이는(단계 601~603).
- 54> 반대로 단계 601에서 읽어온 토큰이 주석문이나 공백이 아니고 응용 프로그램을 위해 정의된 엘리먼트의 시작태그일 경우(단계 604) 엘리먼트의 속성 및 내용을 모두 파싱(단계 605)하여 속성의 끝 즉, 마침태크가 나타날 때까지 읽은 후, 최종적으로 엘리먼트 및 속성의 그래픽 사용자 인터페이스 정보를 저장한다(단계 606~608).

- <55> 그리고, 어휘 분석기(310)에서는 엘리먼트의 내용을 파싱한 후 나머지 토큰을 읽어온다.(단계 609~610)
- <56> 이후, 단계 611에서는 읽어온 토큰들이 일반 사용자 데이터인지를 판단하여 만약, 일반 사용자 데이터이면 단계 612에서 내용의 그래픽 사용자 인터페이스 관련 정보를 저장하고, 일반 사용자 데이터가 아닐 경우에는 단계 613에서와 같이 문자열과 같은 일반 사용자 데이터와 주석문 및 공백, 혹은 엘리먼트의 종료를 알리는 앞서 읽어들이는 태그의 마침태그가 오는지를 판단하게 된다.
- <57> 단계 613에서 읽어 들인 토큰이 마침태그가 오지 않을 경우에는 단계 601부터 반복하고, 마침태그가 올 경우에는 단계 614에서 정의된 시작태그의 마침태그인지를 판단한다.
- <58> 단계 614에서 읽어 온 토큰에서 정의된 마침태그가 오지 않을 경우에는 이를 무시하고(단계 616), 마침태그가 올 경우에는 종료하게 된다.
- <59> 단계 611에서 본 발명은 일반 사용자 데이터일 경우 즉, 문자열과 같이 화면에 보여주어야 할 사용자 데이터가 나오면 관련 정보를 저장하고(단계 612), 현재 엘리먼트의 마침태그를 읽어들이면 엘리먼트 파싱을 종료하며, 그렇지 않고, 다시 응용 프로그램에서 정의한 엘리먼트의 시작태그를 읽어들이게 되면 이를 엘리먼트의 내용으로 간주하여 엘리먼트를 파싱한다.
- <60> 한편, 단계 604에서 본 발명은 응용 프로그램에서 정의되지 않은 엘리먼트의 시작태그를 인식하게 되면, 계속 엘리먼트의 태그와 속성 및 마침태그가 나올 때 까지 읽어들이고, 이들 에 대해 처리하지 않고 초기단계로 리턴한다(단계 615).

【발명의 효과】

- <61> 이상에서 설명한 바와 같이, 본 발명은 통합 파서를 휴대 단말기 탑재할 경우 기존에 구축된 웹 사이트를 그대로 이용할 수 있을 뿐만 아니라, 단말의 응용 프로그램이 필요로 하는 정보만을 추출하는 것이 가능하다.
- <62> 더욱이 본 발명에 의하면, 인터넷 서비스 제공자는 각 단말에 특화된 웹 사이트를 구축하지 않아도 되므로 시간적이나 금전적으로 비용을 절감할 수 있는 효과를 얻을 수 있다.
- <63> 이상에서 설명한 것은 본 발명에 따른 엘리먼트 기반 웹 문서의 파싱 시스템 및 파싱 방법을 설명한 하나의 실시 예에 불과한 것으로써, 본 발명은 상기한 실시 예에 한정되지 않고, 이하의 특허청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변경 실시가 가능한 범위까지 본 발명의 기술적 사상이 미친다고 할 것이다.

【특허청구범위】**【청구항 1】**

웹 문서를 호출하여 휴대 단말에 제공하기 위한 웹 문서 파싱시스템에 있어서,

지원하고자 하는 문서종류에서 필요로 하는 모든 마크업 데이터를 위한 토큰 테이블을 참조하여 마크업과, non-마크업을 기반으로 토큰을 분리하는 어휘 분석기; 및

각 문서의 DTD를 기반으로 콘텐츠 모델을 분석하여 이를 기반으로 각 구문을 분석한 후, 상기 단말의 그래픽 사용자 인터페이스(GUI)를 기반으로 트리 기반의 객체를 생성하는 구문 분석기;를 포함한 것을 특징으로 하는 휴대 단말의 응용에 적용하기 위한 엘리먼트 기반 웹 문서의 파싱시스템.

【청구항 2】

제 1항에 있어서, 상기 어휘 분석기는,

주석이나 공백을 처리하는 주석문 분석기, 마크업 시작태그를 인식하여 토큰을 생성하는 마크업 시작분석기, 속성을 분석하여 토큰을 생성하는 속성 분석기 및 일반 사용자 데이터를 분석하여 토큰을 생성하는 일반 사용자 데이터 분석기를 포함하는 것을 특징으로 하는 엘리먼트 기반 웹 문서의 파싱시스템.

【청구항 3】

제 1항에 있어서, 상기 구문 분석기는,

상기 어휘 분석기에서 생성된 토큰을 바탕으로 해당 문서가 각각의 DTD에 적합하게 작성되었는지 여부를 검사하는 XML 검증부 및 상기 분석된 마크업을 단말의 그래픽 사용자 인터페

이스와 일치시키는 단말 GUI기반 객체 생성부로 이루어진 것을 특징으로 하는 엘리먼트 기반 웹 문서의 파싱시스템.

【청구항 4】

제 1항 내지 제 3항 중 어느 한 항에 있어서, 상기 파싱 시스템은,

HTML, XHTML, MHTML, CHTML, WML, HDML들과 관련된 SGML 또는 XML을 기반으로 작성된 웹 문서를 통합하여 파싱하는 것을 특징으로 하는 엘리먼트 기반 웹 문서의 파싱시스템.

【청구항 5】

제 1항 내지 제 3항 중 어느 한 항에 있어서, 상기 파싱 시스템은,

특정한 휴대 단말에 제한되지 않으며, 각 단말의 사양에 따라 파싱할 엘리먼트의 종류를 선택할 수 있는 것을 특징으로 하는 엘리먼트 기반 웹 문서의 파싱시스템.

【청구항 6】

웹 서버로부터 호출한 웹 문서를 파싱하기 위한 방법에 있어서,

상기 웹 문서로부터 토큰을 읽어와 분석하는 제 1단계;

상기 토큰을 분석한 결과 정의된 시작태그가 아니거나 주석문 또는 공백일 경우 그 토큰을 무시하고, 정의된 시작태그가 읽힐 경우 상기 토큰으로부터 엘리먼트의 속성을 파싱하는 제 2단계;

상기 토큰으로부터 엘리먼트 속성을 파싱하고 엘리먼트의 GUI관련 정보를 저장한 후 엘리먼트의 내용을 파싱하는 제 3단계;

상기 3단계에서 엘리먼트의 내용을 파싱한 결과 일반 사용자 데이터이면 내용의 GUI관련 정보를 저장하고, 일반 사용자 데이터가 아닐 경우에 마침태크가 나올 때까지 데이터를 읽는 제 4단계; 및

상기 제 4단계에서 일반 사용자 데이터가 아닌 상태에서 정의된 시작태그의 마침태크가 나타나면 종료하고, 그렇지 않을 경우 무시하고 리턴하는 제 5단계;를 포함하는 것을 특징으로 하는 휴대 단말에 적용하기 위한 엘리먼트 기반 웹 문서의 파싱방법.

【청구항 7】

제 6항에 있어서, 상기 제 2단계에서,

읽어 온 토큰에 정의된 시작태그가 없을 경우 마침태그가 나올 때까지 계속 읽은 후 그 토큰을 무시하고, 새로운 토큰을 읽어오도록 하는 것을 특징으로 하는 엘리먼트 기반 웹 문서의 파싱방법.

【청구항 8】

웹 서버로부터 호출한 웹 문서를 파싱하기 위한 프로그램이 기록된 기록매체에 있어서,

상기 웹 문서로부터 토큰을 읽어와 분석하는 제 1기능;

상기 토큰을 분석한 결과 정의된 시작태그가 아니거나 주석문 또는 공백일 경우 그 토큰을 무시하고, 정의된 시작태그가 읽힐 경우 상기 토큰으로부터 엘리먼트의 속성을 파싱하는 제 2기능;

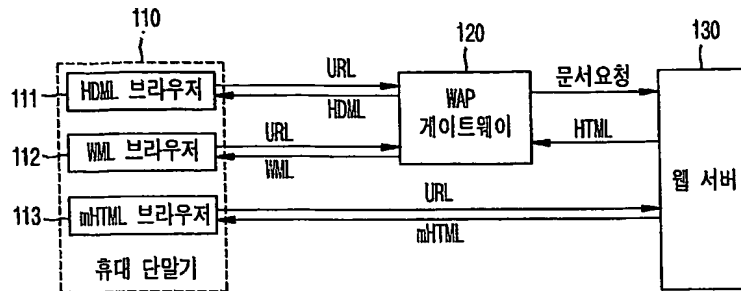
상기 토큰으로부터 엘리먼트 속성을 파싱하고 엘리먼트의 GUI관련 정보를 저장한 후 엘리먼트의 내용을 파싱하는 제 3기능;

상기 3단계에서 엘리먼트의 내용을 파싱한 결과 일반 사용자 데이터이면 내용의 GUI 관련 정보를 저장하고, 일반 사용자 데이터가 아닐 경우에 마침태크가 나올 때까지 데이터를 읽는 제 4기능; 및

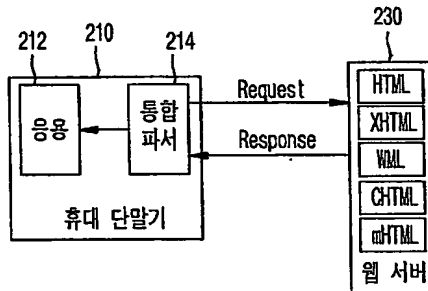
상기 제 4기능에서 일반 사용자 데이터가 아닌 상태에서 정의된 시작태그의 마침태크가 나타나면 종료하고, 그렇지 않을 경우 무시하고 리턴하는 제 5기능;을 실현시키기 위한 프로그램이 기록된 컴퓨터로 읽을 수 있는 기록매체.

【도면】

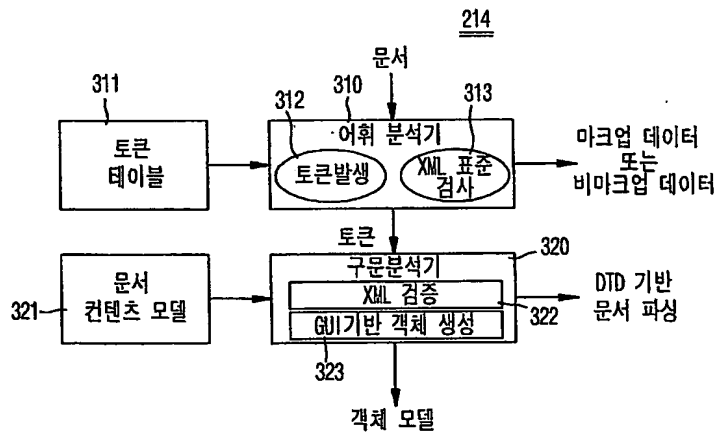
【도 1】



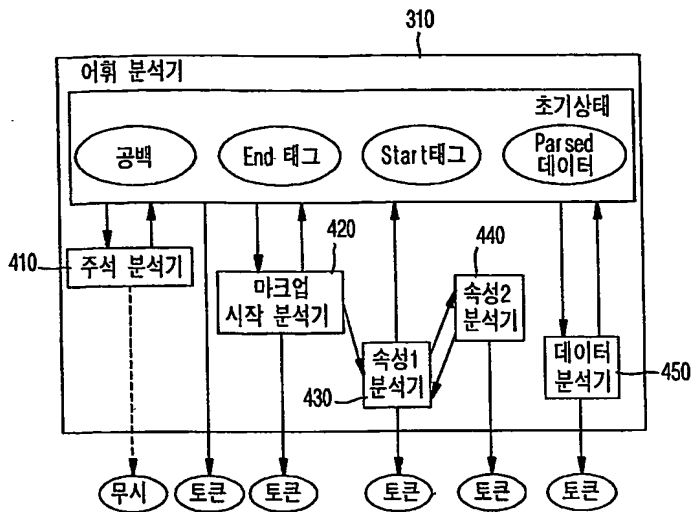
【도 2】



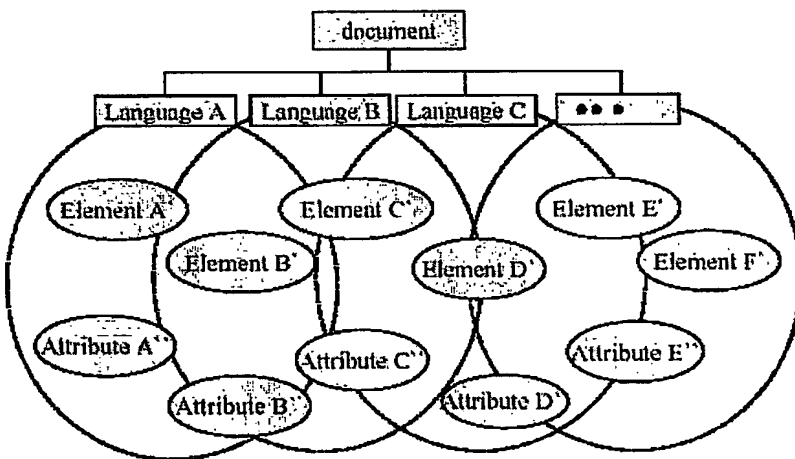
【도 3】



【도 4】



【도 5】



【도 6】

